Read the News, not the Books: Predicting Firms' Financial Health

Shuang (Sophie) Zhai
Ivy College of Business
Iowa State University
szhai@iastate.edu

Zhu (Drew) Zhang
Ivy College of Business
Iowa State University
zhuzhuang@iastate.edu

Abstract

Can one predict a firm's future debt ratio without reading its accounting books? In this paper, we show that mining firm-related events in news articles can effectively predict various quantitative measures of corporate financial health. By exploiting state-of-the-art neural architectures including pseudo event embeddings, LSTMs, and attention mechanisms, the news-powered deep learning models are shown to outperform standard econometric models operating on precise historical data. We also observe prediction performance improvement in multi-task learning settings, i.e., when multiple financial measures are predicted simultaneously. Our predictive models (and byproducts such as corporate embeddings) are poised to benefit various stakeholders, in particular when accounting data is not available.

1 Introduction

Quantitative analysis of corporate financial measures is a well-known technique for stakeholders to exam company financial performance. However, a firm's financial performance is complex and dynamic. Although investors can utilize financial reports to learn inside information about a company, there is a considerable disadvantage of the company financial report release: they are retrospective. It means company financial reports are released a long time after the fact and generally months of delay. Hence investors cannot use financial reports to gain company knowledge on time.

Can stakeholders obtain information about their interested companies before the financial report release? One straightforward approach is to seek information from news articles. News articles are frequently posted online, and most time for free. There are many advantages when investors obtain information from the news. Firstly, when accounting information is not available, news articles are naturally considered as an information source. Investors can use the publicly available information in the news to infer company standing. Secondly, news articles are generally published immediately after an event has happened so readers can capture up-to-date information from the news report. Thirdly, news coverage is often accompanied by experts opinions, and that means readers can observe other alternative views. Fourthly, if investors systematically follow a company's news, the collective information can be triangulated and used to form better-educated investment decisions. Also, stakeholders constantly look for financial opportunities especially when the accounting book is not yet available. So the news is a great information source for them.

Many finance-related works have utilized econometric models to predict a firm's financial measures [Fang and Peress, 2009, Tetlock, 2010, Edmans, 2011]. However, although deep learning models have become increasingly popular in finance applications [Ding et al., 2014, 2015], very little attention has been paid to corporate finance measures prediction. Knowing that predicting corporate financial status in a timely manner will help investors to make better investment decisions, we investigate the

NIPS 2018 Workshop on Challenges and Opportunities for AI in Financial Services: the Impact of Fairness, Explainability, Accuracy, and Privacy, Montréal, Canada. This is a non-archival publication - the authors may submit revisions and extensions of this paper to other publication venues.

important problem of predicting corporate finance measures from news articles in this paper. The high dimensionality of textual data present challenges to traditional econometric models, such as ARIMA, and naturally renders itself to deep learning type of models. We examine various neural architectures for the corporate finance measure prediction problem.

2 Related literature

In the corporate finance literature, many quantitative measures have been used to evaluate a company's financial well-being, such as its capital structure and market value [Brealey et al., 2012]. Every financial measure assesses the company's financial health from a different angle. For instance, capitalization measures the debt component in a company's capital structure. Efficiency measures the level of effectiveness of the company on utilizing its assets and liabilities. Financial soundness and solvency measure a firm's capability to meet its long-term financial obligation. Liquidity measures evaluate if a company can fulfill its short-term financial obligation. Profitability measures assess a firm's capability to receive profits. Valuation measures the company's stock value on the market. By jointly evaluating multiple financial measures from different categories, one can obtain a holistic view of the interested company and get informed about the company's financial status. [1]

Related to our work, researchers have studied how to use textual data to predict stock return [Tetlock, 2011, 2010] and stock price movement [Ding et al., 2014, Boudoukh et al., 2013, Mittermayer and Knolmayer, 2006, Hu et al., 2018]. Researchers have also studied the stock price problem from different data sources, such as messages from stock message board [Antweiler and Frank, 2004], social media [Tetlock, 2007], news stories [Lavrenko et al., 2000], and earning call transcripts [Wang and Hua, 2014]. They have employed various technique to solve the problem, for instance, bag of words, noun phrases, name entities [Schumaker and Chen, 2009], word embeddings [Li et al., 2017], and events [Ding et al., 2015, 2016]. A recent survey [Li et al., 2018] comprehensively examined the relationship between Web media and the stock market. Chakraborty et al. [2016] used a generative model, which was inspired by Latent Dirichlet Allocation (LDA) [Blei et al., 2003] model, to generate event classes from news articles. They combined extracted events with Autoregressive Integrated Moving Average (ARIMA) model for food price prediction.

Our paper also contributes to the growing literature on financial news, reports [Hwang and Kim, 2017], and earning announcement studies [Berkman et al., 2009]. Instead of using sentiment [Das and Chen, 2007, Tetlock et al., 2008, Garcia, 2013], investor moods [Bollen et al., 2011], or economic [Bali et al., 2017] analysis, we use state-of-art machine learning techniques to analyze high-dimensional text data.

3 Related techniques

AutoRegressive Integrated Moving Average (ARIMA) Models. The ARIMA model is the de facto standard for univariate time series analysis in econometrics, and hence our baseline model. ARIMA has an autoregression (AR) component on variable itself, a differencing (I) component if the time series is not stationary, and a moving average (MA) component on error terms [Pankratz, 1983]. In ARIMA(p, d, q) model, p denotes the order of the autoregression (AR) model, d denotes the order of differencing (I), and q denotes the order of the moving average (MA) model.

$$\phi(B)\nabla^d Y_t = \theta(B)\epsilon_t \tag{1}$$

$$\epsilon_t \sim N(0, \sigma^2) \tag{2}$$

where ∇ is the difference operator, $\phi(B) = \phi_p(B) = 1 - \sum_{i=1}^{p} \phi_i B^i$ is the p-order AR operator, $\theta(B) = \theta_j(B) = 1 - \sum_{j=1}^{q} \theta_j B^j$ is the q-order MA operator, and ϵ_t is a white noise process.

Long Short-Term Memory Networks (LSTMs). Long Short-Term Memory (LSTM) is a representative deep learning architecture, and it was first proposed by Hochreiter and Schmidhuber [1997].

[1] Categories are introduced in Wharton Research Data Services (WRDS) https://wrds-web.wharton.upenn.edu/wrds/.

Table 1: Sample variable definition

Category	Variable Name	Code	Formula
Efficiency	*Sales to Working Capital*	sale_nwc	sales/working capital
Valuation	*Price to Earning (diluted, excl.EI)*	pe_exi	price/earnings
Solvency	*Debt Ratio*	de_ratio	total liability/shareholders' equity

There are numerous works have been done in natural language processing by using LSTMs, such as part of speech (POS) tagging, information extraction, syntactic parsing, speech recognition, machine translation [Bahdanau et al., 2014], and question answering.

Bidirectional LSTM (BiLSTM) has not only a forward pass but also a backward pass which can be used to convey later time information backward to earlier time steps.

$$h_t = \sigma(\overrightarrow{h_t}, \overleftarrow{h_t}) \tag{3}$$

Attention Mechanism. Attention mechanism has been explored broadly in recent publications. The intuition for attention technique in natural language processing is to assign higher attention to texts where contains more information for the task on hand. Yang et al. [2016] employed both word-level and sentence-level attentions for document classification task. Wang et al. [2016] proposed an aspect-level attention to capture different sentiments for different aspects in a sentence. Ma et al. [2017] proposed an interactive attention architecture that models the interaction between the context and target for sentiment classification task. More recently, Vaswani et al. [2017] used multi-head attentions alone to solve sequence prediction problems which are traditionally handled by other neural network techniques such as Long Short-Term Memory and Convolution Neural Networks (CNN).

Sentence Encoding. A centerpiece in deep learning models for text mining is semantic representations of words and sentences. Glove [Pennington et al., 2014] vectors are widely used as word embedding. The authors used global matrix factorization and local context window techniques to learn word representations in the vector space. Glove trains word embedding on non-zero elements in word-word co-occurrence matrix and provides a non-zero 300-dimension vector for each word. Word2Vec [Mikolov et al., 2013] is another well-known word embedding method, which includes the Continuous Bag-of-Words (CBOW) and the Skip-gram models. The CBOW model uses the surrounding words in the words window to predict the center word, whereas the skip-gram model uses the center word to predict the surrounding words inside the words window.

While one can trivially combine word embedding to produce a sentence embedding, researchers have studied techniques particularly for sentence encoding, such as SIF [Arora et al., 2016]. SIF is an unsupervised method, and it provides one 300-dimension representation for each sentence. SIF represents each sentence as a weighted average of word embeddings, applies PCA/SVD, and removes the top principal components. Despite the simplicity of SIF, it is shown to achieve better performance than sophisticated methods such as LSTMs on several supervised tasks.

4 Problem definition

Let's formally define the problem as

$$Y^H_{(i)q} = f^M_{j=0}(E_{(i)j}) \tag{4}$$

$$E_{(i)j} = g^{|S|}_{k=0}(S_{(i)jk}) \tag{5}$$

where,

- Y is one of target measures, summarized in Table 1.
- i indexes companies; j indexes time windows; k indexes sentences; q indexes target measures.

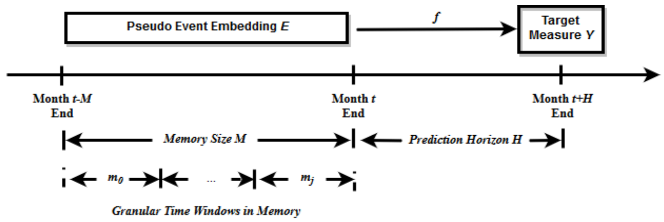


Figure 1: Time structure of the model

- H denotes the prediction horizon, and M denotes the size of memory, both are measured in terms of the number of time windows.
- $E_{(i)j}$ is the pseudo event embedding for company C_i in time window m_j; $S_{(i)jk}$ is the encoding of a sentence k that describes company C_i in time window m_j.
- f is a learned model (function) that maps all pseudo event embeddings in memory to target values.
- g is a function that aggregates multiple sentence embeddings into one pseudo event embedding.

In principle, f and g can be parameterized as any function approximator.

There are broadly speaking two categories of the business news we see. One category is explicitly describing a company's financial measurement change. We list the following sentences as examples.

- *JPMorgan* increased 5.2 percent to $34.98.
- *Wal-Mart* added 1.4 percent to $60.58.
- *Google* gained less than 1 percent to $625.96.

The second category includes what has happened and what is going to happen to the company. The following are several sample sentences, which we treat as pseudo-events (the italicized entities represent the focal companies under consideration). The intuition behind our problem formulation is that semantic signals encoded in such events have predictive power for the focal company's future financial status and well-being.

(a) To describe what has happened:

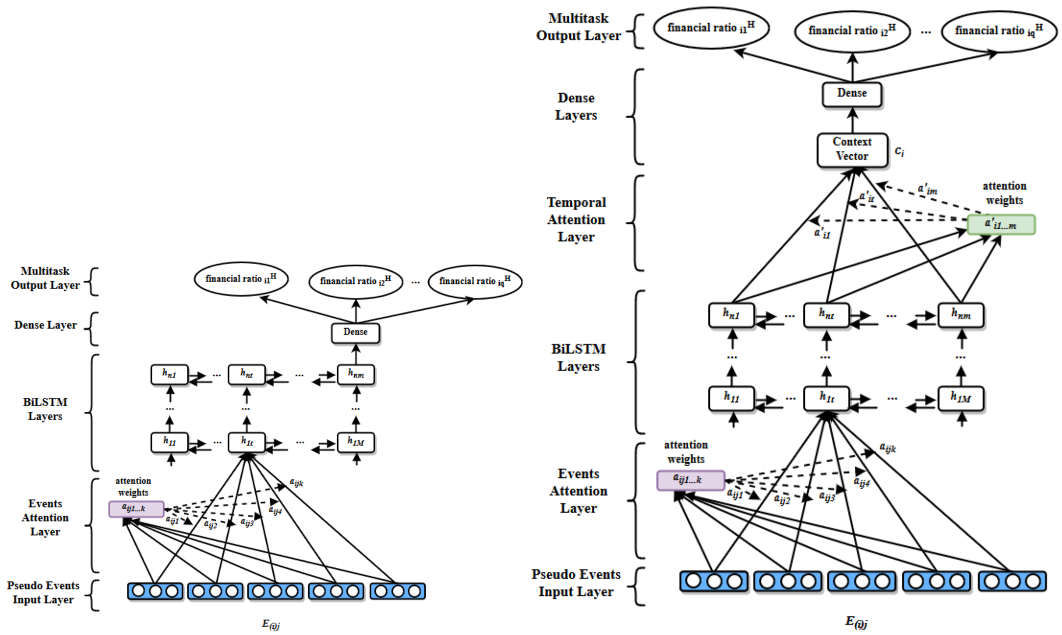
- *Amazon.com Inc.* invested $175 million in LivingSocial in December.
- A group led by Apple Inc. and *Microsoft Corp.* agreed earlier this year to buy Nortel Networks Corp.'s patent portfolio for $4.5 billion.
- Chicago-based *Boeing Co.*, the world's largest aerospace company, said it received 127 orders in August, up from 115 in July.

(b) To describe what is going to happen:

- *AT&T* said that it would expand the rollout of its high-speed wireless technology, called Long-Term Evolution, or LTE, under the T-Mobile agreement.
- *Time Warner Cable* was willing to pay 6 percent more in 2012, according to Witmer, who said MSG was asking for dollars more than any other sports network. [sentence in 2011]
- Now, as rulings start coming in, it might be time for a détente that helps *Apple* maximize the value of its patents, said Kevin Rivette, a managing partner at 3LP Advisors LLC, a firm that advises on intellectual property.

5 Models

Our models are illustrated in Figure 2. We institute various treatments of function g in Equation 5, such as attention mechanism. We use *SIF* sentence embedding method [Arora et al., 2016] to encode our pseudo-events. And within each time window, we compute every pseudo-events embedding and generate the top k vectors according to the L2 norm of all embeddings in the same time window, for each company. In Figure 2a, we implement events attention on the top k pseudo-events for each time window and each company. The weighted sum context vector is used as the semantic representation of the company within the time window. This architecture is good for both single-task prediction and multi-task prediction. We compare the single-task and multi-task performance in later sections. In Figure 2b, we add temporal attention to distinguish information for one company across time. The temporal weighted sum context vector c_i is used in both single-task and multi-task learning predictions as well.



(a) Events attention model

(b) Events+temporal attention model

Figure 2: Events attention and events+temporal attention models

We define our event attention as:

$$\widetilde{h}_{it} = sigmoid(W_{ijk}E_{ijk} + b_{ij}) \tag{6}$$

$$\alpha_{ijk} = softmax(W_{it}\widetilde{h}_{it}) \tag{7}$$

$$h_{it} = \sum_{k=1}^{K} \alpha_{ijk}E_{ijk} \tag{8}$$

To further enhance the model, we also build temporal attentions to weight historical news differently. We define our temporal attention as:

$$\widetilde{c}_i = ELU(W_{im}h_{im} + b_{im}) \tag{9}$$

$$\alpha\prime_{im} = softmax(W_i\widetilde{c}_i) \tag{10}$$

$$c_i = \sum_{m=1}^{M} \alpha\prime_{im}h_{im} \tag{11}$$

6 Experiments

Evaluation Models. We evaluate the following models in our experiments:
(a) **ARIMA**: the baseline.
(b) **X-EvAttn**: events attention model for single task prediction.
(c) **Multi-EvAttn**: events attention model for multi-task prediction.
(d) **X-TempoAttn**: events + temporal attention model for single task prediction.
(e) **Multi-TempoAttn**: events + temporal attention model for multi-task prediction.

Table 2: Dataset statistics

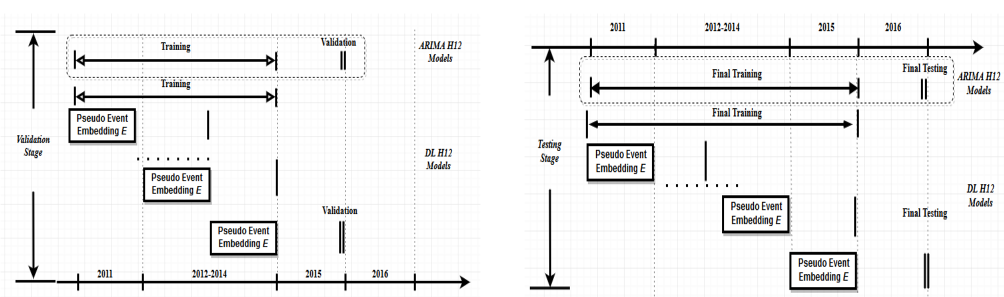
Year	Number of Articles	Number of Sentences
2011	39,696	179,468
2012	46,049	194,074
2013	80,791	398,208
2014	85,577	338,566
2015	93,203	329,571
Total	**345,316**	**1,439,887**

Data. We crawled news articles from `Bloomberg.com` published between the year 2011 and 2015. Our study focuses on the Fortune 1,000 companies, and we use them as our focal companies in the experiments. We collected each company's ticker ID and financial measures from Wharton Research Data Services (WRDS). The dataset statistics are listed in Table 2.

Preprocessing. We extracted each news content and published date and time from our dataset. By consulting a pre-compiled gazetteer of company names, name variants, and abbreviations, we extracted all sentences that have the focal company name or name variants presented. Please note one sentence can be extracted multiple times if there are more than one in-list companies are mentioned in the particular sentence. Finally, we replaced all company names and their variants with a special token 'FOCOMP' (meaning 'focal company'), in order to minimize the noise in the sentence encoding process that caused by company name which contains common English words (e.g., apple, machines, etc.). We use Python NLTK for news content and sentence extractions. After the above process, the vocabulary size of our model is 294,842, and the total number of words is 64,959,753. Also, to tackle the problem of missing values in the company financial ratio sequences and be precise at the same time, we first excluded companies that have more than two missing values, then filled in the missing values by linear interpolation for each company, separately.

Model Training and Evaluation Setup. In our experiments, we define each time window m_j to be a month, and we set memory size M = 12. In our dataset, the median number of pseudo-events per company per month is 5. So we use k=5 in our experiments. We set up the problem as long-term financial measure prediction, as illustrated in Figure 1, hence we set the prediction horizon H = 12.

We use two stages to test our models. First, we use data from the year 2011 to 2014 as training data and predict measures at the end of the year 2015. We use the predicted measures as validation to tune our models. After we confirm the model parameters from validation, we retrain the same model by using all data from the year 2011 to 2015 and evaluate the final prediction at the end of 2016. We illustrate our validation and testing setup in Figure 3.



(a) Validation (b) Evaluation

Figure 3: Validation and evaluation stage scenario

ARIMA model is well-known to be suited for small memory sizes, so we look for our baseline model among a set of small memory size ARIMA models. To be able to find the best performing ARIMA model parameter, we experimented combinations of the following parameter settings: autoregressive term p in $\{1, 2, 3\}$, differencing term d in $\{0, 1\}$, and the moving average term q in $\{0, 1\}$. Among the above ARIMA model variations, we use parameters (p, d, q) from the best performing ARIMA model in the final testing.

We use exponential linear units (ELUs) [Clevert et al., 2015] as activation function in all of our deep learning models. We use Adam [Kingma and Ba, 2014] as the optimizer. And we employ Mean Absolute Percentage Error (MAPE) as the loss function and performance measure.

7 Results

We show our model performance in Table 3.

Table 3: Model performance (MAPE)

Variable	ARIMA	X-EvAttn	Multi-EvAttn	X-TempoAttn	Multi-TempoAttn
sale_nwc	100.55%	57.42%	**54.65%**	55.46%	**53.33%**
pe_exi	100.28%	98.27%	**97.12%**	92.50%	**94.69%**
de_ratio	66.92%	57.59%	**52.39%**	54.91%	**51.32%**

First of all, we can observe the ARIMA results are relatively high, and it suggests us the non-linearity characteristic of the time series. This behavior also indicates the limit of ARIMA models. By using company events information alone, our X-EvAttn model can outperform ARIMA significantly. Furthermore, if we predict multiple tasks simultaneously, we can see the Multi-EvAttn model improves performance on all variables. If we desire to weight news at different points in time, we can add temporal attention to the model. After adding temporal attention, X-TempAttn models outperform X-EvAttn on all variables. Moreover, when we predict multiple measures simultaneously by using both events and temporal attention, we continue to observe model performance improvements. Although not all of the three variables obtained performance gain on Multi-TempoAttn compares to X-TempAttn, it hinted a plausible path when one wants to increase the performance of one or a group of variables. In other words, one variable (pe_exi in this case) helped other variables (sale_nwc and de_ratio in this case) to continue to achieve performance gain while sacrificing its own performance. This approach implies that we can bundle variables together and they can share information with each other during the learning process, and performance advancement can be maximized.

8 Discussion

8.1 Events and temporal attentions visualization

To be able to see what event attention and temporal attention provide us, we visualize the pseudo-events attention weights (in blue) and temporal attention weights (in red) in Figure 4. The highest weighted pseudo-event is also highlighted in each time window. First of all, we can see events attention tends to assign higher weight on the event that will have a longer impact, especially when the highest weight is much higher than other weights within the same period. For instance, the highest attention pseudo-event in month 1 is talking about the focal company filed a patent application. The filing patent event will likely to create a long-term impact on the company financially. And in month 6, the pseudo-event suggests investors sell the focal company stocks. Among the temporal attentions, we can see the model tends to assign higher weights to more recent events. It is intuitive from a human perspective to allocate more attention to more recent events as well.

8.2 Corporate embeddings

A byproduct of our model is corporate embeddings. Our models can generate corporate embedding without knowing which company it is since we replaced all company names with the special token 'FOCOMP'. We use the Fama French 5 industry portfolio to segment companies. It segments companies into 5 industries, including (1) *Consumer* (2) *Manufacturing* (3) *HiTech* (4) *Health*and (5) *Others*: mines, construction, building management, transaction, hotels, business service, entertainment, and finance.

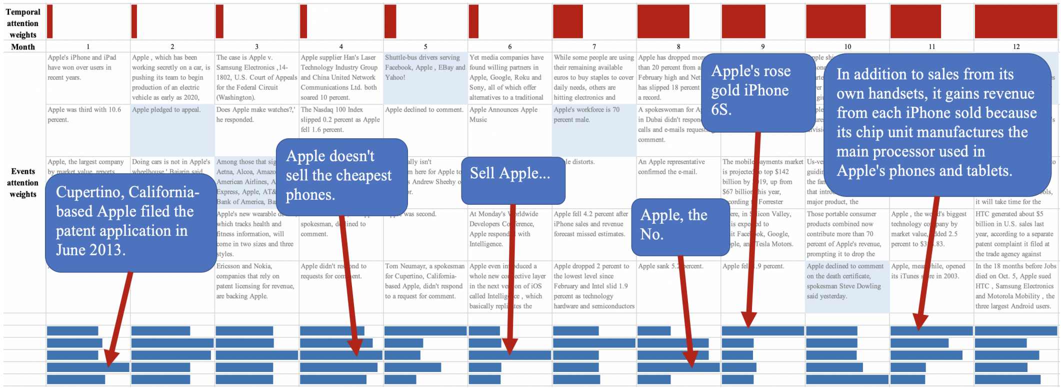
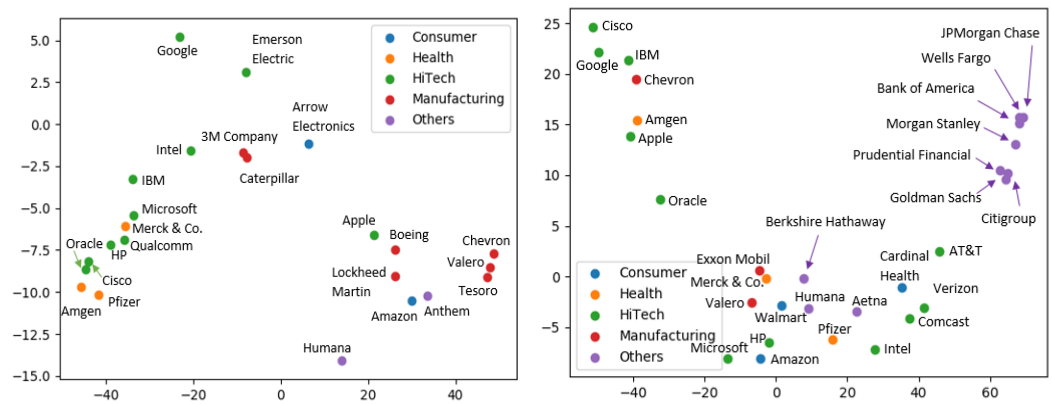


Figure 4: Events and temporal attention visualization



(a) sale_nwc (b) de_ratio

Figure 5: Corporate embeddings encoded by sale_nwc and de_ratio

Surprisingly, our corporate embedding did very well on clustering similar companies by only consuming pseudo-events information from public news. We should also note our corporate embedding is very flexible in investigating company financial well-being from multiple perspectives. The different corporate embeddings encoded by different financial measures show various aspects of the firm's financial health. As we illustrated in Figure 5a, when the corporate embedding is encoded by sale_nwc, we observe Google's sales and working capital structure is very different from other well-known technology companies. The corporate embedding will also be very beneficial when investors are looking for appropriate company peers. Another example is Amazon. Amazon is categorized in *Consumer* industry in Fama French industry portfolio. In Figure 5a, Amazon is far from tech companies. However, in Figure 5b, albeit it's *Consumer* industry background, Amazon's debt structure makes it very like a tech company, such as Microsoft and HP. Another interesting company we see on 5b is Berkshire Hathaway. It stays far away from other financial services firms, such as Wells Fargo and JPMorgan Chase. However, it remains relatively close to insurance companies such as Humana and Aetna. It suggests the company's viewpoint towards company indebtedness.

9 Conclusion

In this paper, we show the effectiveness of firm-related event model on predicting corporate financial measures. The news empowered deep learning models are shown to outperform standard econometric models. We see the performance improvement on multiple measures in multi-task learning. Our predictive model and corporate embedding can also provide meaningful insights for decision making, especially when accounting data is not available.

References

Werner Antweiler and Murray Z Frank. Is all that talk just noise? the information content of internet stock message boards. *The Journal of finance*, 59(3):1259–1294, 2004.

Sanjeev Arora, Yingyu Liang, and Tengyu Ma. A simple but tough-to-beat baseline for sentence embeddings. 2016.

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.

Turan G Bali, Stephen J Brown, and Yi Tang. Is economic uncertainty priced in the cross-section of stock returns? *Journal of Financial Economics*, 126(3):471–489, 2017.

Henk Berkman, Valentin Dimitrov, Prem C Jain, Paul D Koch, and Sheri Tice. Sell on the news: Differences of opinion, short-sales constraints, and returns around earnings announcements. *Journal of Financial Economics*, 92(3):376–399, 2009.

David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.

Johan Bollen, Huina Mao, and Xiaojun Zeng. Twitter mood predicts the stock market. *Journal of computational science*, 2(1):1–8, 2011.

Jacob Boudoukh, Ronen Feldman, Shimon Kogan, and Matthew Richardson. Which news moves stock prices? a textual analysis. Technical report, National Bureau of Economic Research, 2013.

Richard A Brealey, Stewart C Myers, Franklin Allen, and Pitabas Mohanty. *Principles of corporate finance*. Tata McGraw-Hill Education, 2012.

Sunandan Chakraborty, Ashwin Venkataraman, Srikanth Jagabathula, and Lakshminarayanan Subramanian. Predicting socio-economic indicators using news events. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1455–1464. ACM, 2016.

Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. Fast and accurate deep network learning by exponential linear units (elus). *arXiv preprint arXiv:1511.07289*, 2015.

Sanjiv R Das and Mike Y Chen. Yahoo! for amazon: Sentiment extraction from small talk on the web. *Management science*, 53(9):1375–1388, 2007.

Xiao Ding, Yue Zhang, Ting Liu, and Junwen Duan. Using structured events to predict stock price movement: An empirical investigation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1415–1425, 2014.

Xiao Ding, Yue Zhang, Ting Liu, and Junwen Duan. Deep learning for event-driven stock prediction. In *Ijcai*, pages 2327–2333, 2015.

Xiao Ding, Yue Zhang, Ting Liu, and Junwen Duan. Knowledge-driven event embedding for stock prediction. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 2133–2142, 2016.

Alex Edmans. Does the stock market fully value intangibles? employee satisfaction and equity prices. *Journal of Financial economics*, 101(3):621–640, 2011.

Lily Fang and Joel Peress. Media coverage and the cross-section of stock returns. *The Journal of Finance*, 64(5):2023–2052, 2009.

Diego Garcia. Sentiment during recessions. *The Journal of Finance*, 68(3):1267–1300, 2013.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8): 1735–1780, 1997.

Ziniu Hu, Weiqing Liu, Jiang Bian, Xuanzhe Liu, and Tie-Yan Liu. Listening to chaotic whispers: A deep learning framework for news-oriented stock trend prediction. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pages 261–269. ACM, 2018.

Byoung-Hyoun Hwang and Hugh Hoikwang Kim. It pays to write well. *Journal of Financial Economics*, 124(2):373–394, 2017.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Victor Lavrenko, Matt Schmill, Dawn Lawrie, Paul Ogilvie, David Jensen, and James Allan. Mining of concurrent text and time series. In *KDD-2000 Workshop on Text Mining*, volume 2000, pages 37–44, 2000.

Qing Li, Jun Wang, Feng Wang, Ping Li, Ling Liu, and Yuanzhu Chen. The role of social sentiment in stock markets: a view from joint effects of multiple information sources. *Multimedia Tools and Applications*, 76(10):12315–12345, 2017.

Qing Li, Yan Chen, Jun Wang, Yuanzhu Chen, and Hsinchun Chen. Web media and stock markets: A survey and future directions from a big data perspective. *IEEE Transactions on Knowledge and Data Engineering*, 30(2):381–399, 2018.

Dehong Ma, Sujian Li, Xiaodong Zhang, and Houfeng Wang. Interactive attention networks for aspect-level sentiment classification. *arXiv preprint arXiv:1709.00893*, 2017.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

Marc-Andre Mittermayer and Gerhard F Knolmayer. Newscats: A news categorization and trading system. In *Data Mining, 2006. ICDM'06. Sixth International Conference on*, pages 1002–1007. Ieee, 2006.

Alan Pankratz. Forecasting with univariate box-jenkins models: concepts and cases. 1983.

Jeffrey Pennington, Richard Socher, and Christopher Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.

Robert P Schumaker and Hsinchun Chen. Textual analysis of stock market prediction using breaking financial news: The azfin text system. *ACM Transactions on Information Systems (TOIS)*, 27(2): 12, 2009.

Paul C Tetlock. Giving content to investor sentiment: The role of media in the stock market. *The Journal of finance*, 62(3):1139–1168, 2007.

Paul C Tetlock. Does public financial news resolve asymmetric information? *The Review of Financial Studies*, 23(9):3520–3557, 2010.

Paul C Tetlock. All the news that's fit to reprint: Do investors react to stale information? *The Review of Financial Studies*, 24(5):1481–1512, 2011.

Paul C Tetlock, Maytal Saar-Tsechansky, and Sofus Macskassy. More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance*, 63(3):1437–1467, 2008.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, pages 5998–6008, 2017.

William Yang Wang and Zhenhao Hua. A semiparametric gaussian copula regression model for predicting financial risks from earnings calls. In *ACL (1)*, pages 1155–1165, 2014.

Yequan Wang, Minlie Huang, Li Zhao, et al. Attention-based lstm for aspect-level sentiment classification. In *Proceedings of the 2016 conference on empirical methods in natural language processing*, pages 606–615, 2016.

Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. Hierarchical attention networks for document classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1480–1489, 2016.